ROCKY MOUNTAIN
JOURNAL OF MATHEMATICS
Volume 20, Number 2, Spring 1990

ON THE EQUIVALENCE OF THE OPERATOR EQUATIONS $XA + BX = C$ AND $X - p(-B)Xp(A)^{-1} = W$ IN A HILBERT SPACE, p A POLYNOMIAL

TAPAS MAZUMDAR AND DAVID F. MILLER

ABSTRACT. We consider the solution of $(*)$ $XA + BX = C$ for bounded operators A, B, C and X on a Hilbert space, A normal. We establish the existence of a polynomial p and a bounded operator W with the property that the unique solution X of $(*)$ also solves $X - p(-B)Xp(A)^{-1} = W$ uniquely. A known iterative algorithm can be applied to the latter equation to solve $(*)$.

1. Introduction and notations. We know that the equation

$$XA + BX = C \tag{1}$$

with

$$\sigma(A) \cap \sigma(-B) = \phi \tag{2}$$

in which A, B, C are given finite-dimensional matrices of compatible orders, and $\sigma(T)$ is the spectrum of the matrix T (or possibly the operator T), has a unique matrix solution X [**7,9**]. Letting $r(T)$ denote the spectral radius of T, an iterative method to calculate the matrix solution X of the system (1), (2) is obtained *if* we can rewrite (1) in an equivalent form

$$X - UXV = W \tag{3}$$

with

$$r(U)r(V) < 1. \tag{4}$$

When this is possible, the recursion

$$X_{k+1} = U^{2^k} X_k V^{2^k} + X_k, \qquad X_0 = W \tag{5}$$

Received by the editors on November 6, 1987 and, in revised form on March 21, 1988.

can be iterated to convergence, with $X_k \to X$ the unqiue solution of (1). Details of this iterative method are given in [**9**], where the assumption

(6) A and B have only eigenvalues with negative real parts

was imposed to ensure that (4) was satisfied. In [**4**] the assumption (6) was eliminated, and a process for constructing a polynomial p was given so that a solution of (1), (2) is also a solution of (3), (4) if

$$U = p(-B) \text{ and } V = p(A)^{-1}. \tag{7}$$

Polynomials are computer-friendly, and (4), (7) permit computer generated iterates to converge to the solution of the system (1), (2).

The work in [**4**] is restricted to finite dimensions. The purpose of the present article is to generalize some of the results of [**4**] to infinite dimensional Hilbert spaces. We intend to prove that a polynomial p exists so that (3), (4) and (7) are true whenever X is a solution of (1), given (2). As an example, A might be a Hilbert-Schmidt operator [**1, 2**] and B might be a unitary operator [**2, 3**], or B might be an operator such that $\sigma(-B)$ consists of only a finite number of eigenvalues. Left open is the general question of devising a method for constructing such a polynomial p, although a few comments in this direction will be made near the end of §3. Also left unattempted is a generalization of our work to cases where A, B might be differential operators defined on suitably selected Hilbert spaces.

We begin by developing our system of notations rigorously. $\mathcal{H}$ is a separable Hilbert space over the complex numbers $\mathbf{C}$, with norm $||\cdot||$ and innerproduct $((\cdot,\cdot))$. $\mathcal{B}(\mathcal{H})$ is the usual complex Banach space of all bounded linear operators from $\mathcal{H}$ into $\mathcal{H}$, having a norm again denoted by $||\cdot||$ for convenience. We will now have $A, B, C, X \in \mathcal{B}(\mathcal{H})$, and it is in this context that the system (1), (2) will henceforth be viewed. For $t \in \mathbf{R}$, the real numbers, and $T \in \mathcal{B}(\mathcal{H})$, the symbol e^{tT} makes sense as $\sum_{n=0}^{\infty}(tT)^n/n!$ which converges absolutely and uniformly in t over every compact subset of $\mathbf{R}$. Thus $e^{tT} \in \mathcal{B}(\mathcal{H})$. For $g \in \mathcal{H}$, the various needed continuity, differentiability and integrability properties of the map $t \mapsto e^{tT}g : \mathbf{R} \to \mathcal{H}$ will be obviously satisfied.

Let $\mathcal{X}, \mathcal{Y} : \mathbf{R} \to \mathcal{B}(\mathcal{H})$ be two (vector-valued) maps. $\mathcal{Y}$ is called the derivative of $\mathcal{X}$ in the scalar sense, if $((\mathcal{Y}(t)h,k)) = \frac{d}{dt}((\mathcal{X}(t)h,k))$ for all $h, k \in \mathcal{H}$. $\mathcal{Y}$ is called the derivative of $\mathcal{X}$ in the strong (i.e., norm) sense if $||\mathcal{Y}(t) - \frac{\mathcal{X}(t+\Delta t)-\mathcal{X}(t)}{\Delta t}|| \to 0$ as $\Delta t \to 0$. The proof of Theorem 2.1 appearing in the next section will go through, no matter in which sense the derivative $\mathcal{L}'$ of the map $\mathcal{L}$ appearing there is viewed. Similarly the proof will go through independently of how the integrals of the vector-valued functions appearing there are viewed, the strong-Riemann way or the scalar-Riemann way.

2. Main results. We continue with (1) and (2) in the Hilbert space context. In the remainder of this paper, assume that (2) is satisfied. Under this assumption, it is well-known [**5**] that (1) possesses a unique solution. We have the following result.

THEOREM 2.1. *If* $X \in \mathcal{B}(\mathcal{H})$ *solves* (1), *then* X *solves*

$$(Xe^{tA} - e^{-tB}X)h = e^{-tB}\int_0^t e^{\tau B}Ce^{\tau A}h\,d\tau \tag{8}$$

for all $t \in \mathbf{R}$ *and all* $h \in \mathcal{H}$.

PROOF. The proof is patterned after [**4**]. Define $\mathcal{L} : \mathbf{R} \to \mathcal{B}(\mathcal{H})$ by defining $\mathcal{L}(\tau)$ by

$$\mathcal{L}(\tau)h = e^{\tau B}Ce^{\tau A}h \quad \text{for all } h \in \mathcal{H}. \tag{9}$$

Then $\mathcal{L}(0)h = Ch$ and $\mathcal{L}'(\tau)h = B\mathcal{L}(\tau)h + \mathcal{L}(\tau)Ah$. Calculating $\int_0^s \mathcal{L}'(\tau)e^{-sA}hd\tau$ in two ways we see that, for all $s \in \mathbf{R}, h \in \mathcal{H}$,

$$e^{sB}Ch - Ce^{-sA}h = \int_0^s [B\mathcal{L}(\tau) + \mathcal{L}(\tau)A]e^{-sA}h\,d\tau.$$

On rearrangement, this yields

$$e^{-sB}Ce^{-sA}h = Ch - e^{-sB}\int_0^s [B\mathcal{L}(\tau) + \mathcal{L}(\tau)A]e^{-sA}h\,d\tau.$$

So, by (1), (9),

$$\begin{aligned} &e^{-sB}XAe^{-sA}h + e^{-sB}BXe^{-sA}h \\ &\quad = e^{-sB}\mathcal{L}(s)e^{-sA}h \\ &\qquad + \int_0^s [e^{-sB}(-B)\mathcal{L}(\tau)e^{-sA}h + e^{-sB}\mathcal{L}(\tau)(-A)e^{-sA}h]\,d\tau. \end{aligned}$$

Another way of writing this equation is

$$-\frac{d}{ds}(e^{-sB}Xe^{-sA}h) = \frac{d}{ds}\int_0^s e^{-sB}\mathcal{L}(\tau)e^{-sA}h\,d\tau.$$

Integrating over s from 0 to t gives

$$Xh - e^{-tB}Xe^{-tA}h = \int_0^t e^{-tB}\mathcal{L}(\tau)e^{-tA}h\,d\tau$$

which yields (8) as soon as we choose $h = e^{tA}g$. □

COROLLARY 2.2. *If $X \in \mathcal{B}(\mathcal{H})$ solves* (1), *then, for every positive integer n, and every $t \in \mathbf{R}$, we have*

$$XA^n - (-B)^n X = Q_n, \tag{10}$$

where

$$\begin{aligned} Q_n &= (-B)^{n-1}C + Q_{n-1}A \quad (= (-B)Q_{n-1} + CA^{n-1}) \\ &= \sum_{i=1}^{n}(-B)^{n-i}CA^{i-1}. \end{aligned} \tag{11}$$

PROOF. We will show

$$XA^n e^{tA} - (-B)^n e^{-tB}X = (-B)^n e^{-tB}\int_0^t e^{\tau B}Ce^{\tau A}\,d\tau + Q_n e^{tA}. \tag{12}$$

Equation (10) then follows by setting $t = 0$ in (12). We prove (12) by induction. Straightforward differentiation of (8) with respect to t yields (12) for $n = 1$. Here Q_1 is given by (11) with $n = 1$. Straightforward differentiation of (12) with respect to t completes the induction step. □

COROLLARY 2.3. *If $X \in \mathcal{B}(\mathcal{H})$ solves* (1) *then, for every polynomial p, we have*

$$Xp(A) - p(-B)X = W_p, \tag{13}$$

where $W_p = \sum_{i=0}^n a_i Q_i$ whenever $p(z) = \sum_{i=0}^n a_i z^i$. We take $Q_0 = \mathbf{0}$.

THEOREM 2.4. (CONVERSE TO THEOREM 2.1). *If X satisfies* (8) *for t in an open interval around the origin, then X solves* (1).

PROOF. Differentiate (8) with respect to t, and then put $t = 0$. □

LEMMA 2.5. *Assume that*

$$A \text{ is a normal operator } [1, 2, 3] \tag{14}$$

and that

$$\begin{aligned}&\text{the complement of } \sigma(A) \cup \sigma(-B) \text{ in } \mathbf{C}\\ &\text{is a connected subset of } \mathbf{C}.\end{aligned} \tag{15}$$

Then there exists a polynomial p with complex coefficients such that

$$(p(A))^{-1} \text{ exists as an element of } \mathcal{B}(\mathcal{H}) \tag{16}$$

and

$$r(p(A)^{-1}) r(p(-B)) < 1. \tag{17}$$

PROOF. Define $f : \sigma(A) \cup \sigma(-B) \to \mathbf{C}$ by

$$f(z) = \begin{cases} 1 & \text{for all } z \in \sigma(A) \\ \frac{3}{8} & \text{for all } z \in \sigma(-B). \end{cases} \tag{18}$$

We quote the statement of Mergelyan's Theorem from [**6**, p. 386]:

"*If K is a compact set in the plane whose complement is connected, if f is a continuous complex function on K which is holomorphic in the interior of K, and if $\varepsilon > 0$, then there exists a polynomial P such that $|f(z) - P(z)| < \varepsilon$ for all $z \in K$.*"

The theorem is valid (see [**6**]) when the interior of K is empty and when K is disconnected.

We apply Mergelyan's theorem with $K = \sigma(A) \cup \sigma(-B)$, f given by (18), and $\varepsilon = 1/4$. Because of (15) there exists a complex polynomial p such that

$$|f(z) - p(z)| < \frac{1}{4} \quad \text{for all } z \in \sigma(A) \cup \sigma(-B). \tag{19}$$

In particular, by (18), (19),

$$\frac{4}{3} > \frac{1}{|p(z)|} > \frac{4}{5} \quad \text{for all } z \in \sigma(A). \tag{20}$$

Thus $q(z) = (p(z))^{-1}$ is a continuous function on $\sigma(A)$. By the spectral isomorphism property of normal bounded operators [**1, 2**], $q(A) = (p(A))^{-1} \in \mathcal{B}(\mathcal{H})$. Next,

$$r(p(A)^{-1}) = \sup\{|\zeta| : \zeta \in \sigma(p(A)^{-1})\} = \sup\{|p(z)^{-1}| : z \in \sigma(A)\},$$

by the spectral mapping theorem for normal operators [**1**, p. 245]. So, by (20),

$$r(p(A)^{-1}) \leq \frac{4}{3}. \tag{21}$$

Also, $r(p(-B)) = \sup\{|p(z)| : z \in \sigma(-B)\}$ by the spectral mapping theorem for polynomials [**3**, p. 381]. By (18) and (19) we have

$$\frac{1}{8} < |p(z)| < \frac{5}{8} \quad \text{for all } z \in \sigma(-B), \tag{22}$$

so that $r(p(-B)) \leq 5/8$. This, together with (21), yields (17). $\square$

The preceding results can now be combined to yield

Theorem 2.6. *Under the assumptions* (2), (14), (15), *the unique solution X of* (1) *is a solution of the equation*

$$X - p(-B)Xp(A)^{-1} = W \tag{23}$$

for some complex polynomial $p(z) = \sum_{j=0}^{n} a_j z^j$ for which (16), (17) *are satisfied. Here $W = W_p(p(A))^{-1}$, where W_p is given by Corollary* 2.3.

We have now generalized that part of [**4**] which deals with the existence of a suitable polynomial p.

THEOREM 2.7. *In addition to* (2), (14), *let us assume further that*

$$K_1 \cap K_2 = \phi, \ \mathbf{C} \setminus (K_1 \cup K_2) \text{ is connected,} \\ \text{where } K_1 = \sigma(A) \cup \sigma(A^*) \text{ and } K_2 = \sigma(-B) \cup \sigma(-B^*), \tag{24}$$

A^*, B^* being the Hilbert space adjoints of A, B respectively. Then there exists a polynomial $p_{\mathbf{R}}(z)$ with real coefficients for which (16), (17) *are true with p replaced by $p_{\mathbf{R}}$. Furthermore, the unique solution of* (1) *is also a solution of* (23) *with p replaced by $p_{\mathbf{R}}$.*

PROOF. Let $K = K_1 \cup K_2$. Define $f : K \to \mathbf{C}$ by $f(z) = 1$ for all $z \in K_1$, and $f(z) = 3/8$ for all $z \in K_2$. As in the proof of Lemma 2.5, there exists a complex polynomial $p_1(z)$ such that

$$\frac{3}{4} < |p_1(z)| < \frac{5}{4}, \quad \text{for all } z \in K_1, \tag{25a}$$

and

$$\frac{1}{8} < |p_1(z)| < \frac{5}{8}, \quad \text{for all } z \in K_2. \tag{25b}$$

By the fundamental theorem of algebra, we may write

$$p_1(z) = re^{i\theta} \prod_{j=1}^{n} (z - \lambda_j) \tag{26}$$

for some $r > 0, \theta \in \mathbf{R}, \lambda_j \in \mathbf{C}$ for $j = 1, 2, \dots, n, n$ a positive integer. In (26), i is the imaginary unit. Set now

$$p_{\mathbf{R}}(z) = p_1(z) re^{-i\theta} \prod_{j=1}^{n} (z - \overline{\lambda}_j),$$

$\overline{\lambda}_j$ denoting the complex conjugate of λ_j. $p_{\mathbf{R}}(z)$ is a polynomial with real coefficients because of the presence of the multiplicative pair $(z - \lambda_j)(z - \overline{\lambda}_j)$ for every j. Furthermore, for all $z \in \mathbf{C}$,

$$\begin{aligned} |p_{\mathbf{R}}(z)| &= |p_1(z)| r \prod_{j=1}^{n} |z - \overline{\lambda}_j| \\ &= |p_1(z)| r \prod_{j=1}^{n} |\overline{z} - \lambda_j| = |p_1(z)|\ |p_1(\overline{z})| \quad \text{by (26).} \end{aligned}$$

We know that $z \in \sigma(T) \Leftrightarrow \overline{z} \in \sigma(T^*)$ for all $T \in \mathcal{B}(\mathcal{H})$. So, from (25),

$$\frac{9}{16} < |p_{\mathbf{R}}(z)| < \frac{25}{16}, \quad \text{for all } z \in \sigma(A),$$

and

$$\frac{1}{64} < |p_{\mathbf{R}}(z)| < \frac{25}{64}, \quad \text{for all } z \in \sigma(-B),$$

because, for example, $z \in \sigma(A) \Rightarrow \overline{z} \in K_1 \Rightarrow 3/4 < |p_1(\overline{z})| < 5/4$. The rest of the proof is as before (see Lemma 2.5). We have, in particular, $16/9 > 1/|p_{\mathbf{R}}(z)| > 16/25$ for all $z \in \sigma(A), r(p_{\mathbf{R}}(A)^{-1}) \leq 16/9, r(p_{\mathbf{R}}(-B)) \leq 25/64, r(p_{\mathbf{R}}(A)^{-1}) r(p_{\mathbf{R}}(-B)) \leq 25/36 < 1$. □

3. Remarks on applicability of iterative methods.

It is now easily seen that the iterative method (5) given in [**9**, §2] can be applied to solve (23) in view of (17). The question remains whether the solution of (23) arrived at in this way is also the solution of (1). The next theorem answers this question in the affirmative by establishing the equivalence of (1) and (23). For the solution of (1) in a closed integral form, the reader may see [**5**] and the references cited there.

THEOREM 3.1. *Equation* (23) *has a unique solution if p is the polynomial appearing in the proof of Lemma* 2.5 *or if p stands for the polynomial $p_{\mathbf{R}}$ in the proof of Theorem* 2.7 *above.*

PROOF. All that needs to be shown is that $\sigma(p(A)) \cap \sigma(p(-B)) = \phi$ (and then apply [**5**, Corollary 3.3(ii)]). Referring to Lemma 2.5, this is obvious because otherwise we have $\lambda \in \sigma(p(A)) \cap \sigma(p(-B))$ such that

$|\lambda| > 3/4 = 6/8$ by (20), while $|\lambda| < 5/8$ by (22), a contradiction. The proof for $p_{\mathbf{R}}$ as in Theorem 2.7 is analogous. □

Without the aid of the inequalities (20) and (22), the proof of Theorem 3.1 would break down. However, (16) and (17) would then make a partial rescue effort fruitful: A solution of the system (16), (17), (23) differs from a solution of (1) by at most an operator E satisfying $E^2 = \mathbf{0}$. This can be seen as follows.

Let X be a solution of (1) and Y be a solution of (23). Then both X and Y are solutions of (13). Thus, we have

$$\begin{aligned} &(X-Y)A_1 + B_1(X-Y) = \mathbf{0} \\ &\quad \text{with} \;\; A_1 = p(A), \qquad B_1 = -p(-B). \end{aligned} \tag{27}$$

For every positive integer m, z^m is a polynomial in z. With this polynomial apply Corollary 2.3, not to (1) this time, but to (27). We obtain $(X-Y)A_1^m - (-B_1)^m(X-Y) = \mathbf{0}$. Therefore, setting $E = X - Y$, we get $E(p(A))^m = (p(-B))^m E$. Post-multiplying by $(p(A)^{-1})^m E$ and taking norms, we get

$$||E^2|| \leq ||p(-B)^m|| \; ||E|| \; ||(p(A)^{-1})^m|| \; ||E||.$$

Therefore,

$$||E^2||^{\frac{1}{m}} \leq ||p(-B)^m||^{\frac{1}{m}} \; ||(p(A)^{-1})^m||^{\frac{1}{m}} \; ||E||^{\frac{2}{m}}.$$

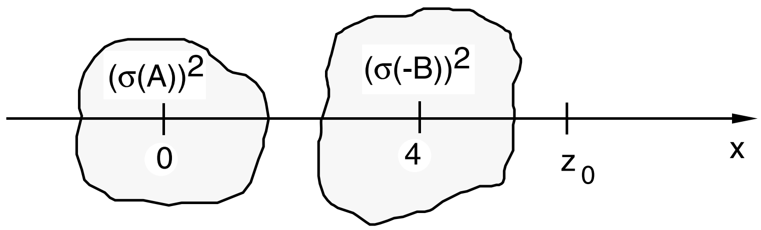
If $E^2 \neq \mathbf{0}$, then we take the limit superior of both sides of this inequality as $m \to \infty$. We obtain

$$1 \leq r(p(-B))r(p(A)^{-1}).$$

This contradicts (17). Hence, we must have $E^2 = \mathbf{0}$.

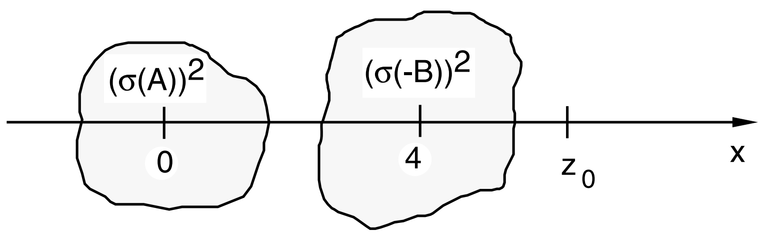
Polynomials satisfying (16), (17) may be easily found in some cases. For example, if $\sigma(A)$ and $\sigma(-B)$ belong to different half-planes, then we may easily choose a point z_0 in the half-plane containing $\sigma(-B)$ such that $\sup_{z\in\sigma(-B)} |z - z_0| < \inf_{z\in\sigma(A)} |z - z_0|$. Then $p(z) = z - z_0$ satisfies (16), (17).

As another example, suppose that $\sigma(A)$ consists of points very close to zero, whereas $\sigma(-B)$ consists of points very close to 2 or -2. If



L is the boundary of the half-planes.

$z \in \sigma(-B)$, then z^2 will be a point close to 4. So $p(z) = z^2 - z_0$ will satisfy (16), (17), z_0 being an appropriately chosen point on the side of $(\sigma(-B))^2 = \{z^2 : z \in \sigma(-B)\}$ away from $(\sigma(A))^2 = \{z^2 : z \in \sigma(A)\}$.



4. Special case when both A, B are self-adjoint. We give below a version of Theorem 2.7 for which an alternate proof is available.

LEMMA 4.1. *If A, B are self-adjoint operators satisfying* (2), *then there exists a polynomial p, with real coefficients, satisfying* (16), (17).

PROOF. Fix a real number $c \geq 2$.

We know that $\sigma(A)$ and $\sigma(-B)$ are compact subsets of $\mathbf{R}$. Fix a compact subset of Δ of $\mathbf{R}$ such that $\Delta \supseteq \sigma(A) \cup \sigma(-B)$.

By virtue of (2), we can apply Urysohn's Lemma [**8**] to infer the existence of a continuous function $f : \Delta \to \mathbf{R}$ such that

$$\begin{aligned} f(z) &= c \quad \text{for all } z \in \sigma(A), \\ f(z) &= \frac{1}{c} \quad \text{for all } z \in \sigma(-B). \end{aligned} \tag{28}$$

By the real Stone-Weierstrass theorem [**7**], there exists a real polynomial p such that

$$|f(z) - p(z)| < \frac{1}{c^2} \quad \text{for all } z \in \Delta. \tag{29}$$

We will now show that this p is the polynomial we are looking for.

From (28) and (29), we have

$$\frac{c^2}{c^3 - 1} > \frac{1}{p(z)} > \frac{c^2}{c^3 + 1} > 0, \quad \text{for all } z \in \sigma(A). \tag{30}$$

Thus, $q(z) = (p(z))^{-1}$ is a continuous function on $\sigma(A)$. By the spectral isomorphism property of self-adjoint bounded operators [1, 2, 3], $q(A) = p(A)^{-1} \in \mathcal{B}(\mathcal{H})$. Next,

$$\begin{aligned} r(p(A)^{-1}) &= \sup\{|\zeta| : \zeta \in \sigma(p(A)^{-1})\} \\ &= \sup\{|p(z)^{-1}| : z \in \sigma(A)\} \end{aligned}$$

by the spectral mapping theorem. So, by (30),

$$r(p(A)^{-1}) \leq \frac{c^2}{c^3 - 1}. \tag{31}$$

Also, $r(p(-B)) = \sup\{|p(z)| : z \in \sigma(-B)\}$ by the spectral mapping theorem for polynomials [**3**, p. 381]. By (28) and (29), we have

$$\frac{c-1}{c^2} < p(z) < \frac{c+1}{2} \quad \text{for all } z \in \sigma(-B),$$

and so $r(p(-B)) \leq (c+1)/c^2$. This, together with (31), yields (17). $\square$

Corresponding versions of Theorems 2.6 and 3.1, and their proofs are now easy to formulate.

REFERENCES

1. J.B. Conway, *A course in functional analysis,* Graduate Texts in Math. **96**, Springer-Verlag, New York, 1985.

2. N.J. Dunford and J.T. Schwartz, *Linear pperators,* Part II, Interscience Publishers, New York, 1963.

3. E. Kreyszig, *Introductory functional analysis with applications,* John Wiley, New York, 1978.

4. D.F. Miller, *A general theory for the iterative solution of* $XA + BX + C = 0$, Linear Algebra Appl. **105** (1988), 131–137.

5. M. Rosenblum, *On the operator equation* $BX - XA = Q$, Duke Math. J. **23** (1956), 263–269.

6. W. Rudin, *Real and complex analysis,* McGraw-Hill, 1966.

7. D.E. Rutherford, *On the solution of the matrix equation* $AX + XB = C$, Akademie van Wetenschappen, Amsterdam, Proc. Amer. Math. Soc. **35** (1932), 54–59.

8. G.F. Simmons, *Introduction to topology and modern analysis,* McGraw-Hill, New York, 1963.

9. R.A. Smith, *Matrix equation* $XA + BX = C$, SIAM J. Appl. Math. **16** (1968), 198–201.

Department of Mathematics and Statistics, Wright State University, Dayton, OH 45435